A New Personalized Recommendation Technique Based on the Modified TOPSIS Method

Guan-Dao Yang[*]

School of Economics
Shanghai University of Finance and Economics
Shanghai 200433, China
E-mail: yangguandao@gmail.com

Lu Sun[*]

Department of Applied Mathematics
Shanghai University of Finance and Economics
Shanghai 200433, China
E-mail: sunlu52@yahoo.com.cn

Abstract — **Personalized recommendation service helping users to target the interesting information from the excessive information set has been widely concerned. In this paper, we firstly propose a new method named Modified TOPSIS Method utilizing the Improved Gray Correlation Analysis Method. Then, we present a new personalized recommendation technique based on the Modified TOPSIS Method. Finally, the verification method utilizing Spearman's Rank Correlation Coefficient demonstrates that our new personalized recommendation technique is efficient.**

Keywords - Personalized Recommendation Technique; Improved Gray Correlation Analysis; Modified TOPSIS Method

I. INTRODUCTION

In such an Information Age, it appeals to be increasingly difficult for individuals to discover the useful information from the overwhelming amount of information available. Personalized recommendation services have been considered as a feasible solution to target proper information adapted to users' preferences and a lot of research efforts have been focused on developing highly reliable personalized recommendation techniques.

In general, there are mainly three types of personalized recommendation system, content-based filtering systems, collaborative filtering systems and data mining systems. Content-based recommendation is the development of collaborative filtering systems. It doesn't need the assessment opinions from the users, which is too subjective and may limited by the insufficiency of users' assessment information. Based the choices the users made before, it can calculate the similarity among users, and then put forward recommendation.

Base on the content-based filtering systems, our new personalized recommendation technique firstly requires users to choose the indices satisfying the customized needs of users. In the second place, after gathering the data sets related to the selected indices, the ranking result calculated by the Modified TOPSIS Method will be eventually presented to users.

II. BACKGROUND

A. The Original TOPSIS Method

The Original TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) Method is a ranking method

based on the principle that the chosen alternative should be as close to the ideal solution, rather than the negative-ideal solution, as possible.

In recent years, as a comprehensive and objective decision analysis method in multi-attribute decision-making problem, the Original TOPSIS Method [1]-[4] has been widely applied in evaluating the performance of indicator-based model, which has no restriction of sample size and data distribution pattern.

Unfortunately, it's undeniable that the selected indices always have some certain relationships with the model that we have structured. Additionally, the change of each index may lead to the internal change of the interactions between indices. All defects mentioned above may contribute to the deviation, which negatively affects the effectiveness of information recommendation.

Nevertheless, the Original TOPSIS Method is failed to take the weights of indices into account. Consequently, the evaluation results calculated by the Original TOPSIS Method are essentially accord with the physical circumstances.

B. The Original Gray Correlation Analysis

In an attempt to reflect the actual impact of each index, some researchers decide to weight. Thus, many methods mainly categorized as subjective and objective have been proposed or subsequently improved. However, those existing weighting methods are still open to question for the reason that the subjective weighting method largely influenced by various subjective factors, like the experience and the research field of experts, will lead to a biased result with ease, whereas the objective weighting method utilizing quantitative technique in Math and Statistics requires a large numbers of data following a normal distribution, which is, of course, difficult in collection, calculation and application.

In order to solve the aforementioned problems, the weighting method utilizing the Gray Correlation Analysis [5]-[6] has been presented.

In the recent years, the Gray Correlation Analysis Method has been widely used in the fields of geology, ecology, economic and so forth. It is an effective method to examine the relationship between statistic data by judging the

---

[*] These authors contribute equally to this work

interrelationship between the two indices based on the extent of their links.

Nevertheless, the Gray Correlation Analysis Method chiefly relies on the traditional technique of relating degree proposed by Julong Deng (1986) has been proven to the problematic. For instance, firstly, the calculated correlation coefficients don't satisfy the properties of uniqueness, symmetry and comparability. Secondly, different resolution ratios result in different correlations. Thirdly, it can merely reflect positive correlation rather than negative correlation. At last, the resolution ratio is generally set as 0.5, which will lead to $r > 0.3333$, therefore, fail to reflect the real situations.

III. THE IMPROVED GRAY CORRELATION ANALYSIS

To resolve the problem of the Original Gray Correlation Analysis mentioned previously, in this section, we introduce but slightly improve the Computation Model of Gray Correlation Degree proposed by Wang Jianqiang (2000) [7].

Denote the selected n indices of indicator-based model by $I_1 \sim I_n$ respectively.

Step 1: Set sequences $X_1 \sim X_n$ according to $I_1 \sim I_n$ respectively at different points $t_1 \sim t_m$, with the raw data of indicator-based model.

What is slightly different from the Computation Model of Gray Correlation Degree is that when calculate the weights, the indices should be adjusted according to the evaluation system they are applied in. If the evaluation system requires higher-better indices, the lower-better ones $X_j (j \in J^-)$ should be turned into higher-better by their own reciprocal value, $\tilde{X}_j = \frac{1}{X_j}$. Finally calculate the weights of $\tilde{X}_j$.

$$\left\{ \tilde{X}_j \mid \tilde{X}_j = X_j, j \in J^+; \tilde{X}_j = \frac{1}{X_j}, j \in J^- \right\} \quad (1)$$

Where the set J^+ means the indices are higher better, and the set J^- means the indices are lower better.

$$X_j = \left\{ x_j(t_1), x_j(t_2), \cdots, x_j(t_m) \right\}, j = 1, 2, \cdots, n \quad (2)$$

Step 2: Calculate D_i

$$D_j = \sum_{i=2}^{m} \left| x_j(t_i) - x_j(t_{i-1}) \right|, j = 1, 2, \cdots, n \quad (3)$$

Step 3: Calculate the increment sequence.

$$\Delta y_j(t_i) = x_j(t_i) - x_j(t_{i-1}),$$
$$j = 1, 2, \cdots n, i = 2, 3, \cdots m \quad (4)$$

Step 4: Calculate the correlation coefficient $\xi_{kj}(t_i)$ and correlation R_j, and then rank the correlation.

The correlation coefficient formula between indices i and j is as follows:

$$\xi_{kj}(t_i) = \frac{1}{\left(1 + \left\| \frac{\Delta y_k(t_i)}{D_k} \right| - \left| \frac{\Delta y_j(t_i)}{D_j} \right\| \right)\left(1 + \left| \frac{\min(D_k, D_j)}{\max(D_k, D_j)} \right| \right)} \quad (5)$$
$$k \in [1, n], j \in [1, n], i \in [1, m]$$

And the correlation formula is as follows:

$$R_j = \frac{1}{n} \sum_{k=1}^{n} \xi_{kj}, \quad j \in [1, n] \quad (6)$$

Therefore, we can conclude the ranked correlation sequence according the value of R_j.

Step 5: Normalize the correlation by the formula as follows:

$$\alpha_j = \frac{R_j}{\sum_{j=1}^{n} R_j} \times 100\%, \quad j \in [1, n] \quad (7)$$

Thus, we can conclude the difference between the impacts of each index.

From the analysis above, we can conclude that the weights of indices utilizing the Improved Gray Correlation Analysis obtain the following properties:

- *Symmetry, that is $\xi_{kj} = \xi_{jk}$*. The Correlation Coefficient Table utilizing the Improved Gray Correlation Analysis is symmetric, however, is asymmetric if utilizing the Original Gray Correlation Analysis.

- *Uniqueness.* We can obtain a unique set of weights of indices.

- *Coincidence.* It means that the correlation between two sequences with a smaller slope of curve must be larger than that with a sharper slope of curve.

IV. THE MODIFIED TOPSIS METHOD

In the previous section, we suggested that the Original TOPSIS Method should take the weights of indices into account, so that the actual impact of each index could be precisely identified. Therefore, in this paper, we improve the Original TOPSIS Method utilizing the aforementioned Improved Gray Correlation Analysis.

We introduce weight of each index calculated by the Improved Gray Correlation Analysis to the Original TOPSIS Method for improvement.

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \alpha_j (Z_{ij} - Z^+)^2}, \, i \in [1, m] \qquad (8)$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \alpha_j (Z_{ij} - Z^-)^2}, \, i \in [1, m] \qquad (9)$$

Where, α_j denotes the weight of each index calculated by Gray Correlation Analysis.

By eliminating the mistakes caused by indices which are essentially measured by the same factors, it is useful to improve the evaluation system of indicator-based model.

Compared with the Original TOPSIS Method, the Modified TOPSIS Method is more objective. It is able to avoid the mistakes caused by the indices which are essentially measured by the same factors, without which they would wrongly magnify the impacts growing in the same direction on the evaluation system.

V. THE NEW PERSONALIZED RECOMMENDATION TECHNIQUE UTILIZING THE MODIFIED TOPSIS METHOD

Step 1: Establish a decision matrix $\left(\tilde{X}_{ij} \right)_{m \times n}$, which is a higher-better matrix.

$$\tilde{X}_{ij} = \begin{cases} X_j(t_i), j \in J^+ \\ \dfrac{1}{X_j(t_i)}, j \in J^- \end{cases} \qquad (10)$$

Where the set J^+ means the indices are higher better, and the set J^- means the indices are lower better.

Step 2: Normalize the higher-better matrix.

The normalized value Z_{ij} is calculated as

$$Z_{ij} = \frac{\tilde{X}_{ij}}{\sqrt{\sum_{i=1}^{m} (\tilde{X}_{ij})^2}}, \quad i \in [1, m]; j \in [1, n] \qquad (11)$$

Then calculate Z^+ and Z^- respectively

$$Z_j^+ = \max_i (Z_{ij}), \quad Z_j^- = \min_i (Z_{ij}) \qquad (12)$$

$$Z^+ = (Z_1^+, Z_2^+, \ldots, Z_n^+) \quad Z^- = (Z_1^-, Z_2^-, \ldots, Z_n^-),$$
$$i \in [1, m]; j \in [1, n] \qquad (13)$$

Step 3: Calculate the dispersion of Z^+ and Z^-.

$$D_i^+ = \sqrt{\sum_{j=1}^{n} \alpha_j (Z_{ij} - Z^+)^2}, \, i \in [1, m]$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \alpha_j (Z_{ij} - Z^-)^2}, \, i \in [1, m]$$

Step 4: Calculate C_i and rank it in descending order.

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}, \, i \in [1, m] \qquad (14)$$

Where the index value of C_i is between 0 and 1. The larger the index value is, the better the performance is.

VI. NUMERICAL EXAMPLE

In this section, as an illustration, we apply our new personalized recommendation technique in hospital beds arrangement, which is a common thorny problem in the healthcare industry. The verification is based on the data of the performance of hospital beds arrangement, which is from the Question B for undergraduate of 2009 China Undergraduate Mathematical Contest in Modeling.

A. The Application in Hospital Beds Arrangement

According to the 2008 China Health Statistics Yearbook published by Ministry of Health of P.R China, we select five following indices, which are currently widely used in the medical industry.

[8] I_1: Weekly Standardized Bed Utilization Rate

$= \dfrac{\text{The Actual Value of Weekly Standardized Bed Utilization Rate}}{\text{The Standard Value of Weekly Standardized Bed Utilization Rate}}$

[9] I_2: Weekly Standardized Bed Turnover Ratio

$= \dfrac{\text{The Actual Value of Weekly Standardized Bed Turnover Ratio}}{\text{The Standard Value of Weekly Standardized Bed Turnover Ratio}}$

I_3: Average Days of Stay of Weekly Recoverd Patients

$= \dfrac{\text{The Total Stay Days of Weekly Recoverd Patients}}{\text{The Sum of Weekly Recoverd Patients}}$

[10] I_4: Average Days of Waiting of Weekly Patients

$= \dfrac{\text{Days of Waiting of Weekly Patients}}{\text{The Sum of Weekly Patients}}$

I_5: The Rate of CD Type $= \dfrac{\text{The Sum of Acute Patients}}{\text{The Sum of Patients}}$

Step 1: Calculate the weights of indices utilizing Gray Correlation Analysis.

First, we get the correlation coefficient table, which satisfies symmetry and coincidence. (See Table 1.)

Secondly, calculate the weights of indices utilizing Gray Correlation Analysis. (See Table 2.)

α_2 has the largest influence to the indicator-based model, followed by α_1 α_5 α_3, the last comes α_4.

Step 2: Structure the higher-better matrix. (See Table 3.)

In Table 2., the raw data of lower-better indicators are changed to higher-better ones.

Step 3: Nondimensionalize the higher-better matrix.

Step 4: Determine Z^+ and Z^-.

According to the matrix calculated in Step 3, we can get Z^+ and Z^- as follows.

$$Z^+ = (1.0240 \quad 0.0285 \quad 0.1667 \quad 0.7916 \quad 0.8571)$$

$$Z^- = (0.0853 \quad 0.0004 \quad 0.0916 \quad 0.0827 \quad 0)$$

Step 5: Calculate the dispersion of Z+ and Z- and make a rank of C_i. (See Table 4.)

Table 4. illustrates that the 1st week ranks first among these weeks, followed by the 9th week, and meanwhile the second week ranks last.

B. The Verification Utilizing Spearman's Rank Correlation Coefficient

In order to verify our Modified TOPSIS Method, we introduce the Spearman's Rank Correlation Coefficient to compare the two evaluating systems based on the Original TOPSIS Method and the Modified TOPSIS Method respectively. The Spearman's Rank Correlation Coefficient [11] is introduced to assess how well the relationship between the two methods can be described utilizing a monotonic function by testing whether the observed value of ρ is significantly different from one.

Step 1: Compute the dispersion of Z+ and Z- using the Original TOPSIS Method and make a rank of C_i. (See Table 5.)

Step 2: Calculate the Spearman's Rank Correlation Coefficient.

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad \in [0,1]$$

$$x_i = (C_i)_{Modified\ TOPSIS\ Method} , \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad (15)$$

$$y_i = (C_i)_{Original\ TOPSIS\ Method} , \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Step 3: Analysis.

Table 4. illustrates that by the means of Modified TOPSIS Method, the first week ranks first among these weeks, followed by the ninth week, which is different from the rank in Table 5.

The Spearman's Rank Correlation Coefficient we get is 0.9753, which means our Modified TOPSIS Method is slightly different from the Original TOPSIS Method. However, since more factors are considered in Modified TOPSIS Method, it is safe to draw the conclusion that our Modified TOPSIS Method is improved.

VII. CONCLUSION

This paper proposes a new personalized recommendation technique utilizing the Modified TOPSIS Method. Firstly, the indices satisfying the customized needs of users should be obtained. In the second place, the weights of indices will be calculated by the Improved Gray Correlation Analysis. At last, the sorted information calculated by the Modified TOPSIS Method will be eventually presented to users.

Compared with the other personalized recommendation techniques, our technique interacts with users directly. Besides, the recommend information meets the personalized needs of users, precisely and timely.

REFERENCES

[1] Ping Song, "The Comprehensive Evaluation of the Hospital Beds Usage Utilizing TOPSIS Method (in Chinese)", Chongqing Medical Journal , Chongqing Institute of Medicine, Chongqing, China, vol. 32, pp. 455-457, April 2003.

[2] Y.-J. Lai, T.-Y. Liu and C.-L. Hwang, "TOPSIS for MODM", European Journal of Operational Research 76 (3), pp. 486-500, 1994.

[3] T.-Ch. Chu, "Facility Location Selection Using Fuzzy TOPSIS under Group Decisions", International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems 10 (6), 687-701, 2002.

[4] S. Mahmoodzadeh, J. Shahrabi, M. Pariazar, and M. S. Zaeri, "Project Selection by Using Fuzzy AHP and TOPSIS Technique", International Journal of Human and Social Sciences, p4-p5, 2007.

[5] Shifang Hou, Zengwen Yue, and Jianguo Xu, "The Analysis on the Factors Affecting the Arrangement of Hospital Beds Utilizing Gray Correlation Analysis (in Chinese)", Chinese Journal of Health Statistics , China Medical University, Shenyang, China, vol. 24, pp. 621-623, Dec 2007.

[6] H. Deng, C.-H. Yeh and R.J. Willis, "Inter-company Comparison Using Modified TOPSIS with Objective Weights", Computers & Operations Research 27 (10), 963-974, (2000).

[7] Wang Jianqiang, "Discussion on the Computation Model of Gray Correlation Degree", Mathematical Theory And Application, Vol.20 No.1, Mar,2000.

[8] Pinggen Wang, Yunsuo Gao, " The Research of the Hospital Beds Arragement in Large General Hospitals (in Chinese)", Chapter 2.4, 2006.

[9] E. Ravi Kiran, K.Vijaya, "Utilization of Beds in A Tertiary Care Hospital", Journal of the Academy of Hospital Administration Vol. 15, No. 2 (July 2003, Dec 2003).

[10] Bar-dayan Yaron ,"Waiting Time is a Major Predictor of Patient Satisfaction in a Primary Military Clinic", Military Medicine, Oct 2002

[11] Wikipedia, Spearman's rank correlation coefficient, http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient, Jan 2010.

AUTHORS PROFILE

Guan-Dao Yang, presently a senior student of the School of Economics, Shanghai University of Finance and Economics, Shanghai, China. Areas of interests include applied mathematics, econometrics and computer science.

Lu Sun, presently a senior student of the Department of Applied Mathematics, Shanghai University of Finance and Economics, Shanghai, China. Areas of interests include applied mathematics and computing science.

Table 1. Correlation Coefficient

	α_1	α_2	α_3	α_4	α_5
α_1	1	1.918548	1.71555	0.963194	1.124417
α_2	1.918548	1	1.034636	1.218497	1.869969
α_3	1.71555	1.034636	1	1.100687	1.69721
α_4	0.963194	1.218497	1.100687	1	0.872475
α_5	1.124417	1.869969	1.69721	0.872475	1

Table 2. Weights of indices

α_1	α_2	α_3	α_4	α_5
20.9854%	21.9843%	20.4434%	16.0936%	20.4933%

Table 3. Higher-better Matrix

0.0853	0.0004	0.16667	0.088689	0.8571
0.4267	0.0022	0.15152	0.084084	0.0986
0.8818	0.0088	0.14388	0.10032	0.2037
1.024	0.0285	0.10586	0.088236	0.1129
0.7111	0.0219	0.1002	0.086495	0.0833
0.8818	0.0272	0.094368	0.082739	0.1273
0.7822	0.0241	0.095986	0.092846	0.3333
0.8818	0.0272	0.09158	0.13964	0.3333
0.4125	0.0127	0.098977	0.79164	0

Table 4. Result of Modified TOPSIS Method

Week	First	Second	Third	Fourth	Fifth	Sixth	Seventh	Eighth	Ninth
D_j^+	0.099815	0.14289	0.1094	0.11401	0.12272	0.11452	0.088724	0.080621	0.091446
D_j^-	0.07658	0.0064175	0.022135	0.044583	0.023329	0.037253	0.03859	0.046931	0.064591
C_i	0.43414	0.04298	0.16828	0.28111	0.15973	0.24545	0.30311	0.36794	0.41395
Ranking	1	9	7	5	8	6	4	3	2

Table 5. Result of Original TOPSIS Method

Week	First	Second	Third	Fourth	Fifth	Sixth	Seventh	Eighth	Ninth
D_j^+	0.5525	0.7660	0.6030	0.6321	0.6743	0.6357	0.5074	0.4584	0.4428
D_j^-	0.3737	0.0310	0.1058	0.2074	0.1082	0.1730	0.1814	0.2207	0.3903

C_i	0.4035	0.0389	0.1493	0.2471	0.1383	0.2139	0.2634	0.3249	0.4685
Ranking	2	9	7	5	8	6	4	3	1